LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-TR-855235

Using Grover's search protocol for the best qubit pair selection

V. I. Geyko, I. Joseph

September 30, 2023

Disclaimer

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

Using Grover's search protocol to select the best qubit pairs

V. I. Geyko and I. Joseph
Lawrence Livermore National Laboratory, Livermore, California, 94550, USA

(Dated: 29 January 2024)

This research represents a continuation of our investigation of the Rigetti quantum platform as part of the Quantum Leap for Fusion Energy Sciences project. We evaluate the performance of the new Aspen-11, Aspen-M-2, Aspen-M-3 quantum processing units (QPU) through the application of the Grover's search algorithm and the validation of single gate fidelities. The performance of the new QPUs is compared to the older Aspen-7 device, and it is shown that qubit pair selection plays a key role in the optimization process. Additionally, we delve into the examination of coherent and decoherent errors associated with native gates. To optimize our approach, we have developed several relatively inexpensive hardware protocols aimed at facilitating the selection of the most suitable qubit pairs. These protocols involve running various circuits on the hardware and assessing the overall performance of the tested qubit pairs. Through these protocols, we have demonstrated that the quality of qubit pairs on a single chip can exhibit significant variations.

I. INTRODUCTION

The recent advancements in quantum computing technology have sparked significant interest in the field of quantum simulations, where these devices have demonstrated their ability to perform complex quantum calculations. The promising concept of "quantum advantage" has generated enthusiasm, leading to extensive research and development efforts by government and private industry laboratories globally. Notably, a substantial focus of these developments has been on scaling up the total number of qubits in quantum computing devices. For instance, within a span of just 6 years, IBM-Q managed to elevate the number of qubits in their superconducting devices from the initial 5 qubits, as seen in the "IBM Q Experience" introduced on May 4, 2016, to 433 qubits in the "IBM Osprey", unveiled on November 9, 2022. This rapid growth in qubit count, while impressive, has brought forth new challenges related to hardware issues, including crosstalk between qubits and the complexity of error-prone qubit control. Consequently, the current state of quantum hardware remains significantly constrained by the absence of fault-tolerant error correction. To address this limitation, an examination of hardware performance has been undertaken, comparing the capabilities of various quantum devices across different Rigetti lattices, spanning from the retired Aspen-7 to the contemporary Aspen-M-3. The results of this investigation are summarized and discussed in the present work.

In the initial stages of our research, we employed a custom-designed 3-level variant of Grover's search[1] (GS) algorithm as a specialized testing tool. This allowed us to benchmark and compare the performance of state-of-the-art quantum platforms, including LLNL QuDIT, IBM-Q, and Rigetti. The motivation for using a 3-level version instead of a conventional 4-level (also called "2-qubit") version is that the transmon of the LLNL QuDIT was capable of doing 3-level control only. Since IBM-Q and Rigetti platforms are based on regular 2-level qubits, all 3-level gates represented as 3×3 unitary matrices were embedded in a 4×4 matrix realized on a pair of adjacent qubits. The 3-level version of the GS algorithm is in fact more complex than the 4-level one. The complexity emerges from the fact that every block of the 3-level GS algorithm is decomposed by the compiler[2] to a long sequence of native gates, where each sequence typically includes 2 or 3 two-qubit entanglement gates, such as the CZ or XY gates. Unlike the 3-level version, the traditional 4-level version of the algorithm includes moderately simple blocks that contain no more than a single entanglement gate only.

When the Rigetti platform was tested in 2020, we discovered a number of hardware issues that dramatically affected the overall performance. They were described in detail in the previous report[3], and we briefly cover them here to present a comprehensive understanding of the issues that are observed.

- *Readout fidelity.* It was shown that fidelity of the state preparation and measurement (SPAM) procedures varied noticeably depending on the time elapsed from the last calibration, aka "retune". Specifically, the readout fidelity exhibited a value of approximately 0.94 immediately following a retune, contrasting with a diminished value of 0.86 observed several hours later. Consequently, to mitigate fidelity loss attributed to SPAM, it was imperative to conduct hardware tests and, notably, production runs only subsequent to a retune.

- *Abnormally large variance.* The system occasionally exhibited spontaneous large amplitude fluctuations in the measured state population. An additional instance of erratic behavior occurred when the state of one or more qubits underwent continuous and rapid temporal drifting, rendering all measurements unreliable. Upon detection, it was recommended to remove such data points to ensure the preservation of reliable data.

- *Decoherent errors and rapid or unexpectedly fast*

fidelity decay. The system's decoherence time T_1 was investigated by assessing the decay rate of an excited state subjected to multiple applications of the CZ gate. Notably, the observed rate was found to be considerably higher than the one estimated based on the T_1 time. These findings formed the foundation for the CZ protocol introduced in this work, further elaborated upon in Sec. (VI).

- *Coherent errors*. Unlike decoherent errors, coherent errors (also recognized in the literaturee as *systematic errors*) don't directly induce fidelity decay, but their cumulative impact could be substantial if they aggregate coherently. For instance, it has been demonstrated that an additional phase accumulates with repeated applications of the RX gate. This phase correction is qubit-dependent and, in theory, could be rectified by applying an additional correction each time the gate is invoked. This matter is revisited and explored in detail in Sec. (V).

As these issues were not well-understood or even identified in 2020, the decision was made to employ a simple 4-level version of the GS algorithm as a testing tool. Alternatively, the focus was directed towards benchmarking individual gates. Ideally, gate set tomography[4] (GST) or process tomography[5] of each native gate is needed to comprehensively characterize the system's behavior. However, due to resource constraints and time limitations, only a relatively small number of tests could be conducted.

In the present report, we extend our ongoing efforts to test and benchmark Rigetti's hardware, specifically the newer QPUs Aspen-11, Aspen-M-2, and Aspen-M-3. It was confirmed that the overall stability of the new QPUs is noticeably better than that of the old Aspen-4/7. No significant qubit drifts or other forms of spontaneous abnormal behavior were detected, and in case of smaller-scale or rare similar issues, the corrupted data points could be safely removed from the statistics. Despite the improved stability, the performance of the the GS algorithm on the new hardware was considerably below expectations, thus, the research shifted towards benchmarking of the individual program constituents and native gates in order to reveal the reasons behind the observed errors. As a result, it was discovered that pairs on the chip behave dramatically differently, and therefore selection of the best qubit pairs is the key step to improve fidelity of any algorithm.

The report is organized as follows. In Sec. (II) we review the GS algorithm. The measurement methodology is introduced in Sec. (III). A comparison of the new data to the old data from Aspen-7 is done in Sec. (IV). Tests of coherent errors of single qubit RX/RZ gates are reported in Sec. (V) as well as possible approaches to correction of these errors by additional over-rotation. The CZ and GS protocols are introduced and the results analyzed in Sec. (VI) and Sec. (VII) respectively. We summarize our results in Sec. (VIII).

II. GROVER'S SEARCH ALGORITHM

A. Theoretical background

Consider a quantum system with N levels, where N is chosen to be either 3 or 4 based on specific requirements and benchmarks conducted. A quantum state is denoted as $|\psi\rangle$, and the basis decomposition reads as

$$|\psi\rangle = \sum_{k=0}^{N-1} \alpha_k |k\rangle, \tag{1}$$

where $|k\rangle$ is the basis vector, such that the only k-th state is occupied, with the standard normalization condition applied

$$\sum_{k=0}^{N-1} |\alpha_k^2| = 1. \tag{2}$$

A 3-level system can be realized on a 2-qubit platform by using three states only ($|00\rangle$, $|01\rangle$, $|10\rangle$). A unitary matrix of a corresponding 3-level gate is of dimensions 3×3, and it is embedded in a 4×4 matrix of a 2-qubit system. Ideally, the unused state should remain unpopulated throughout quantum simulations (which is not perfectly the case in practice) and therefore it should be kept intact by the gates. The choice of the unused states is arbitrary, however in this work, we always assume it is $|3\rangle$, since the matrix structure is simple block-diagonal with a unity in the last cell.

Grover's search quantum circuit consists of the 3 main elements: Superposition gate S, Oracle U_ω, and Grover's diffusion operator U_s. The superposition gate is needed to spread the state vector "evenly" among all the states the search algorithm is operating on, assuming that the initial state is the ground state $|0\rangle$. The superposition state $|s\rangle$ reads as

$$|s\rangle = \frac{1}{N} \sum_{k=0}^{N-1} |k\rangle. \tag{3}$$

For the 4-level GS algorithm, the superposition gate is implemented as a 2-qubit Hadamard gate

$$S_4 = H_2 = H_0 \otimes H_1, \tag{4}$$

with the corresponding unitary matrix

$$S_4 = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \tag{5}$$

For the 3-level system, the S-gate is the discrete Fourier transform[6] with the following matrix representation

$$S_3 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & e^{i\delta} & e^{-i\delta} & 0 \\ 1 & e^{-i\delta} & e^{i\delta} & 0 \\ 0 & 0 & 0 & \sqrt{3} \end{bmatrix} \tag{6}$$

where $\delta = 2\pi/3$, and i is the imaginary unit. It is worth noting that the S_4 gate can also be implemented as the 4-level discrete Fourier transform. However, this choice increases gate complexity and leads to a decrease in gate fidelity.

The Oracle gate can be realized by using either Boolean or phase methods[7]. In this work, we use the latter since the Boolean method requires an ancilla qubit, which brings extra complexity to the system and makes it harder to characterize the hardware performance. The Oracle performs a phase flip of the marked state ω, namely $U_\omega|\psi\rangle = -|\psi\rangle$ for $\langle\psi|\omega\rangle = 1$, which yields to

$$\begin{cases} U_\omega|\psi\rangle = |\psi\rangle, & \text{for } \langle\psi|\omega\rangle = 0; \\ U_\omega|\psi\rangle = -|\psi\rangle, & \text{for } \langle\psi|\omega\rangle = 1. \end{cases} \tag{7}$$

The matrix representation is trivial: it is a unit matrix with -1 instead of 1 at the position (ω, ω).

The Grover's diffusion gate reflects the state vector $|\psi\rangle$ across the superposition vector $|s\rangle$

$$U_s = 2|s\rangle\langle s| - I. \tag{8}$$

The matrix representation is quite dense and takes the following forms for the 3-level system

$$U_{s3} = \frac{1}{3}\begin{bmatrix} -1 & 2 & 2 & 0 \\ 2 & -1 & 2 & 0 \\ 2 & 2 & -1 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix} \tag{9}$$

and for the 4-level system

$$U_{s4} = \frac{1}{2}\begin{bmatrix} -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{bmatrix} \tag{10}$$

The circuit for the Grover's algorithm is schematically illustrated in Fig. (1). Initially, the system is in the ground state $|0\rangle$, and then it is transformed to the superposition state $|s\rangle$ by applying the S gate. Then, the block of U_ω and U_s is applied $n_{\mathtt{it}}$ times. At the end of the algorithm, measurements of all the qubit states are performed. The probability to find the system in the marked Oracle state ω reaches the maximum (very close to unity) at some certain number of Grover's iterations $n_{\mathtt{it}}$. For $N \gg 1$, the number of iterations scales as $n_{\mathtt{it}} \sim \pi\sqrt{N}/4$ yielding a quadratic speed-up[1] over the classical search algorithm that scales as $n_{\mathtt{it}} \sim N$. In the general case, one can prove using straightforward algebra[7] that the population density of the marked state oscillates as

$$p_\omega = \sin^2(n_{\mathtt{it}}\theta + \kappa), \tag{11}$$

where

$$\begin{aligned} \sin(\theta) &= \frac{2\sqrt{N-1}}{N}, \\ \sin(\kappa) &= \frac{1}{\sqrt{N}}. \end{aligned} \tag{12}$$

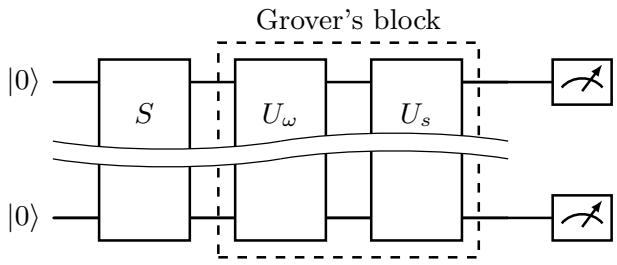

FIG. 1. Grover's search circuit. The superposition gate S transforms the ground state $\psi = |0\rangle$ to the superposition state $|s\rangle$. The block of two gates (Oracle U_ω and Grover's diffusion U_s) is then repeated $n_{\mathtt{it}}$ times. Finally, the the measurements of the quantum state are performed.

The objective of this work is to use Grover's algorithm as a diagnostic tool to assess the performance of Rigetti's hardware. Consequently, we executed the algorithm on various Rigetti quantum processing units (QPUs) and compared the readout values of the Oracle states to the analytical ones for different numbers of Grover's iterations $n_{\mathtt{it}}$. While any number of iterations can be utilized, certain points are of particular significance. These points are when the values of p_ω are close unity, highlighted in Fig. (2) with the green color. This selection is made for the following rationale: assuming that both analytical and measured values of p_ω are around a fraction of unity, for example, 0.3, it becomes challenging to ascertain whether the system is performing well, resulting in a good match, or if the system is nearly decoherent, and the match is merely a coincidental outcome. On the other hand, driving the population of a state to unity is challenging to achieve relying solely on pure noise, decoherence, or dephasing. Hence, the fidelity of the quantum system can be characterized as the deviation of the measured population from the predicted unit value. Based on this approach, the rate of the fidelity decay can be estimated from the observation of the p_ω decay as a function of $n_{\mathtt{it}}$.

As follows from Eq. (11), the set of "suitable" Grover's iterations for the 4-level system is $n_{\mathtt{it}} = \{1, 4, 7, 10...\}$ where $p_\omega = 1$. In the case of the 3-level system, p_ω is never exactly equal to unity. However, there are points $n_{\mathtt{it}} = \{1, 6, 11, ...\}$ where p_ω is closer to unity than typical measurement errors are. It is worth mentioning that in the case of $N = 3$, there are also points of interest where p_ω is very close to zero. The reason is that the total number of states is small; therefore, in a fully decoherent system, one expects to obtain $p_\omega = 1/3$, which is noticeably far from zero. This fact makes these points attractive and they are highlighted with the blue color in Fig. (2).

One example of such a performance test was conducted in 2020 using the old Aspen-7-2Q-C QPU. A 3-level version of the Grover's algorithm was tested on qubit pair (0, 1). The results of that test for Oracle marked states

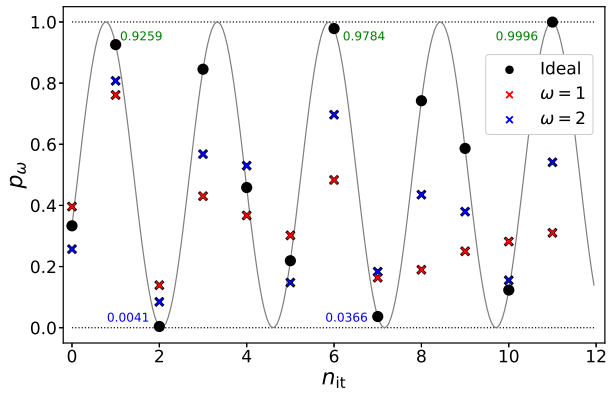

FIG. 2. Probability to find the system in the Oracle marked states $|1\rangle$ (red crosses) and $|2\rangle$ (blue crosses) as a function of the number of Grover's iterations $n_{\mathtt{it}}$. Results collected in 2020 from Aspen-7 QPU are compared to analytical theory (black solid dots). Every point is obtained as an average over 10000 shots. Specific iterations that are useful for determining fidelity are labeled with the green color for p_ω close to unity and with the blue color for p_ω close to zero.

$\omega = \{1, 2\}$ are shown in Fig. (2). The system becomes decoherent relatively quickly, so that the only reliable points of interest are $n_{\mathtt{it}} = \{1, 2\}$. There is also a good match at $n_{\mathtt{it}} = 10$ for $\omega = 2$, but that was probably a coincidence, taking into account how poorly the system behaved at $n_{\mathtt{it}} = 11$. Furthermore, even the probability of the superposition gate (which corresponds to $n_{\mathtt{it}} = 0$) was withing a noticeable spread from 0.25 to 0.4 making the reliability of the other steps questionable. These issues motivated the previous technical report[3] and, in part, the present work.

B. Native gate decomposition and circuit complexity

Every block of the Grover's algorithm is decomposed into a set of native gates by the Rigetti compiler[2]. These gates are $RZ(\phi)$ and $RX(m\pi/2)$ for single qubit operations, where ϕ is an arbitrary angle and m is an integer, and the two-qubit CZ for entanglement (the new XY gate is beyond the scope of the present research). The complexity of the decomposition varies dramatically depending on the gate type. For example, decomposition of the two-qubit Hadamard gate contains only 8 single-qubit gates, 4 gates for each qubit $H = RZ(\pi)RX(\pi/2)RZ(\pi/2)RX(-\pi/2)$, while the decomposition of the 3-level superposition gate contains more than 35 native gates and includes 3 CZ gates.

To facilitate an equitable comparison between the 3- and 4-level GS algorithms, we introduce a metric based on the count of CZ gates for each constituent element in the circuit. This metric ignores errors associated with single-qubit gates and SPAM. Given that two-qubit gates generally exhibit lower fidelity compared to single-qubit gates, the influence of CZ gates is predominant. As the total number of CZ gates becomes sufficiently large, the impact of SPAM errors becomes negligible in the overall assessment. The gate count for both versions of the Grover's algorithm is provided in Table (I). Based on these numbers, one can find the total number of the CZ gates in a circuit with $n_{\mathtt{it}}$ iterations.

GS version \ Gate	S	U_ω	U_s
GS4L	0	1	1
GS3L	3	1	2

TABLE I. CZ gate count of different building blocks of the Grover's algorithm for 3- and 4-level versions.

$$\begin{cases} N_{4L} = 0 + (1+1) \cdot n_{\mathtt{it}} = 2n_{\mathtt{it}}, \\ N_{3L} = 3 + (1+2) \cdot n_{\mathtt{it}} = 3(n_{\mathtt{it}} + 1). \end{cases} \tag{13}$$

III. MEASUREMENT METHODOLOGY

The population of a state $|\omega\rangle$ is measured as an average over N_s independent shots of the same circuit, where each measurements yields to 0 or 1 qubit state. Assuming the mean value is p_ω, the statistical standard deviation is $\Delta_p = \sqrt{p_\omega(1-p_\omega)/N_s}$, which scales as $1/\sqrt{N_s}$. This expression can be different if p_ω is too close to either zero or unity, yet this issue was never observed in any of our tests. For a typical value $p_\omega \sim 2/3$, $\Delta_p = 0.0047$ for $N_s = 10^4$ and $\Delta_p = 0.015$ for $N_s = 10^3$. At the early stages of our research, we used $N_s = 10^4$ to minimize the impact from statistical errors. In later tests, this number was reduced to $N_s = 10^3$ for two main reasons. First, we were limited in resources and available QPU time, therefore it was beneficial to obtain more runs rather than small improvements in precision. Second, it was observed that the statistical spread of the data due to hardware variability was usually noticeably greater than the that one of the limited statistical sampling. In this scenario, $N_s = 10^3$ was more than required in most cases. To mitigate the hardware issues (briefly mentioned in the introduction), we use data batching, i.e. repeating the same measurements n_b times. The mean value is then found as an average over all samples, i.e. by averaging over $n_b N_s$ shots, and the new standard deviation Δ_b is introduced as

$$\Delta_b^2 = \frac{1}{(n_b - 1)} \sum_{j=1}^{n_b} (p_j - p_\omega)^2. \tag{14}$$

If $\Delta_b \geq \Delta_p$, it means that statistical errors are not dominant, and the main source of errors is the hardware. If $\Delta_b \gg \Delta_p$, the system is encountering significant hardware issues, and the data points collected during such instances should be excluded from the statistical analysis.

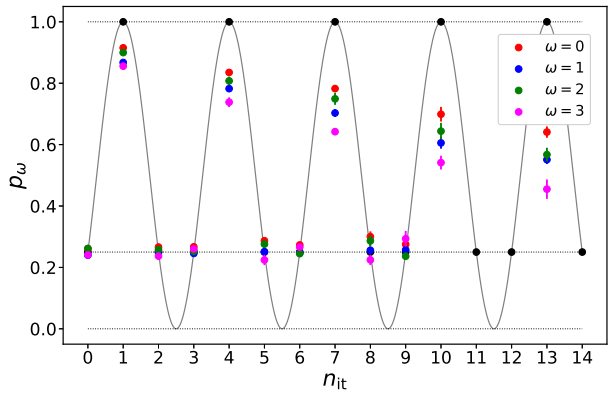

FIG. 3. 4-Level GS algorithm implemented on the Aspen-7 QPU in 2020. Colors correspond to the state $|\omega\rangle$, marked by the Oracle: $\omega = [0, 1, 2, 3]$ →[red, green, blue, magenta]. Solid black dots show analytical values of the desired state population.

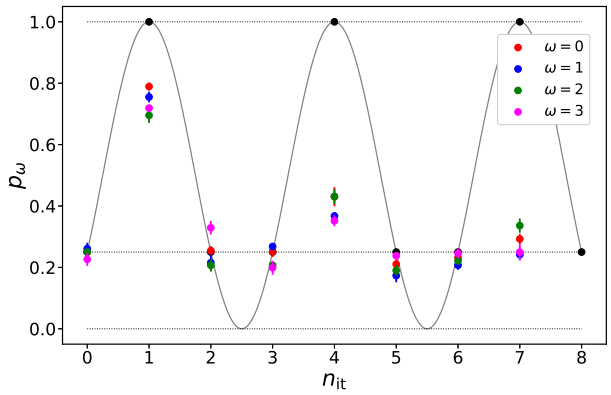

FIG. 4. 4-Level GS algorithm implemented on Aspen-11. The qubit pair is (0, 1). Colors correspond to the state $|\omega\rangle$, marked by the Oracle: $\omega = [0, 1, 2, 3]$ →[red, green, blue, magenta]. Solid black dots show analytical values of the desired state population.

While it is generally good practice to introduce some error mitigation corrections when collecting data, we decided to intentionally avoid this because the goal of the present work is to benchmark the raw performance of the hardware. To explain it further, imagine studying errors of a quantum circuit that are typically comprised of coherent and decoherent errors of the gates as well as SPAM errors. The inclination might be to employ error mitigation techniques, particularly for addressing SPAM errors. However, due to the fact that the nature of the errors is drastically varying across the qubits on a chip, implementing such techniques could inadvertently exacerbate the situation. This complexity introduces challenges in distinguishing the root causes of gate imperfections, potentially complicating the identification of underlying issues.

All test circuits are composed with the PRAGMA PRESERVE BLOCK[8] flag that disables compiler optimization and preserves the original sequence of the gates in a circuit. Furthermore, to mitigate additional errors linked to the variable fidelity of the native gate decomposition of the Grover's search building blocks, we maintained a consistent decomposition of all gates across all algorithm runs. The importance of this aspect was highlighted in previous work[3], emphasizing that an inappropriate choice of decomposition can significantly compromise the interpretation of the results.

An alternative method could involve averaging over random compilations, preventing coherent errors from accumulating and negatively impacting the quantum circuit's performance. However, this approach comes with a tradeoff: the coherent errors are virtually replaced by random-walk type decoherent errors that we have no means of control over. This makes it significantly more challenging to characterize the performance of the QPU, thus, the random compilation approach falls outside the scope of our present work.

IV. COMPARISON OF THE OLD AND NEW HARDWARE PERFORMANCE

The initial series of tests were conducted on the Aspen-11 (QPU). For the default qubit pair on Aspen-11, qubits labeled 0 and 1, which are adjacent on the chip, were chosen. In contrast, the Aspen-7-2Q-C QPU had only two qubits, which eliminated other options for qubit pair selection. Since performance of the 3-level GS algorithm led to rapid decoherence, it was decided to use the simpler 4-level version. Figures (3) and (4) demonstrate the difference of the performance of Aspen-7 and Aspen-11 QPUs. In both cases, the 4-level Grover's search algorithm was implemented for all possible Oracle marked states $\omega = \{0, 1, 2, 3\}$. Every point was obtained according to the data batching procedure described in Sec. (III) with $N_s = 10^4$ and $n_b = 5$. Vertical lines in the figures correspond to error bars defined in Eq. (14). In both cases, data spreads with $\pm\Delta_b$ seem to be significantly smaller than the deviation of the measured results from the analytical theory. In the case of Aspen-7, a considerable decay of fidelity occurs after approximately 13 Grover's iterations, so that there is 0.5 probability to find the system in the desired state. As anticipated, when the Oracle marks the ground state with $\omega = 0$, the algorithm performs noticeably better compared to the $\omega = 3$ state. In the case of Aspen-11, the system becomes decoherent after 7 iterations for any Oracle marked state. The tests of Aspen-11 were repeated several times on different days in order to avoid possible retune issues or accidental spontaneous hardware failures.

The primary factor contributing to the subpar performance of the Aspen-11 was the selection of an unsuitable qubit pair for testing. The *quality* of the qubit pairs can vary drastically, hence the overall stability and fidelity decay as well. There could be various reasons for different behavior of the qubits on a chip, including qubit

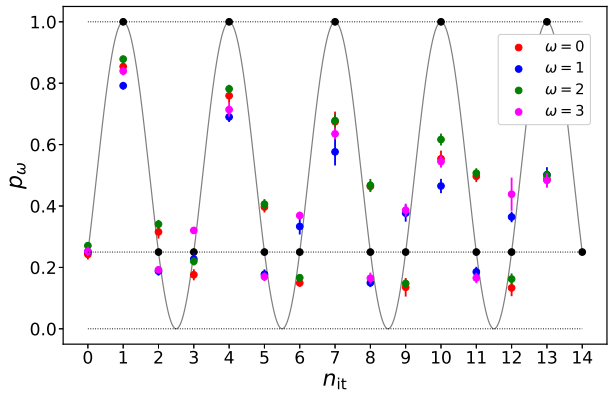

FIG. 5. 4-Level GS algorithm implemented on Aspen-11 in 2022. The qubit pair is (11, 26), one of the best according to the CZ protocol. Colors correspond to the state $|\omega\rangle$, marked by the Oracle: $\omega = [0, 1, 2, 3]$ →[red, green, blue, magenta]. Solid black dots show analytical values of the desired state population.

fabrication quality, control and readout issues, crosstalk, etc., yet further investigation of this issue is out of the scope of the present report. Instead we focus on developing a suitable tool for selection of the best pairs via using either the CZ or GS protocols that are described in detail in Sec. (VI) and Sec. (VII) respectively. These protocols allow one to select a qubit pair of much better quality than average, yielding higher fidelity of the simulations. For comparison, the performance of the 4-level GS algorithm realized on the qubit pair (11, 26) is shown in Fig. (5), and it is markedly better that the one from the pair (0, 1). While the fidelity decay rate from the pair (11, 26) is similar to what we had observed with Aspen-7, the overall behavior of the system is more chaotic and unstable, namely, larger error bars indicate the presence of hardware issues and the growing spread of the measured population at the points $n_{\text{it}} = \{2, 3, 5, 6, 8, 9\}$ from an analytical value of 0.25 indicate the presence of coherent errors. The latter can be quantified by fitting p_ω with the following function that models the decaying oscillations of the state population.

$$p_\omega = \frac{1}{N} + \left[\sin^2(n_{\text{it}}(\theta + \phi) + \theta_0) - \frac{1}{N}\right] e^{-\alpha n_{\text{it}}} \quad (15)$$

Here, N is the number of states (in this case $N = 4$), $\theta = \pi/3$ and $\theta_0 = \pi/6$ are the phase step and the initial phase shift respectively according to Eq. (12). Ideally, the decay coefficient α and the phase shift ϕ are both equal to zero. While both devices had roughly similar decay rates, with $\alpha \sim 0.05 - 0.1$, Aspen-11 suffers from 5-10x larger coherent errors, as determined by $\phi \sim 2 - 4{\times}10^{-2}$ instead of $1 - 9{\times}10^{-3}$.

The latter is also confirmed in Fig. (6). In both cases, the analytical expression from Eq. (15) matches the measured data quite closely, including points at $n_{\text{it}} = \{6, 8, 11\}$ where the phase shift and deviation from the ideal analytical solution are mostly pronounced. The nature of the observed coherent errors is not fully understood. It can be speculated that it is due to crosstalk between qubits on the chip, however, this statement needs additional tests and verification that are beyond the scope of this report.

Oracle marked state ω	Aspen-7-2Q-C		Aspen-11-38Q	
	α	ϕ	α	ϕ
0	0.051804	-0.001742	0.065111	-0.042664
1	0.075157	-0.000261	0.105317	0.0283371
2	0.065053	-0.004931	0.059408	-0.040325
3	0.098371	0.0087434	0.083112	0.0341913

TABLE II. Parameters of the nonlinear fits for the measured data from Aspen-7 and Aspen-11 with decaying function from Eq. (15) for all 4 Oracle marked states $\omega = \{0, 1, 2, 3\}$.

V. RX/RZ GATE TESTS

Following the procedure from our previous work[3], we tested single qubit gates for over/under-rotation. To that end, we employed a simplified version of the *robust phase estimation*[9,10] method. The general formulation of this method allows one to determine both over-rotation angle as well as the angle between the axis of rotation and the $\hat{x}$ axis for the RX gate. In the present research, we focus our efforts on the over-rotation angle only. The set of tests included three thrusts. First, we tested the native gate $RX(\pi/2)$ by repeating it multiple times and measuring accumulation of the phase error. Second, we tested the $RX(\pi/4)$ rotation, which is implemented as a sequence of RX and RZ gates, namely

$$RX\left(\frac{\pi}{4}\right) = \quad (16)$$
$$= RZ\left(\frac{\pi}{2}\right) RX\left(\frac{\pi}{2}\right) RZ\left(\frac{\pi}{4}\right) RX\left(\frac{\pi}{2}\right) RZ\left(\frac{\pi}{2}\right).$$

Notice that the $\pi/4$ rotation is in fact performed by the $RZ(\pi/4)$ gate in the middle of the sequence, as the other two RX/RZ gates only serve the role of state preparation. A noticeable phase shift was detected after around 100 gate repetitions, and it was estimated using a decaying fitting function, similar to Eq. (15)

$$p_\omega = \cos^2(n_{\text{it}}(\theta + \phi) + \theta_0) e^{-\alpha n_{\text{it}}}, \quad (17)$$

where θ_0 and θ are parameters determined by the initial state number (0 and 1) and the rotation angle, and α and ϕ are fitting parameters to determine the decay rate and the angle of over-rotation. Finally, we corrected the over-rotation of the $RX(\pi/4)$ gate by adjusting the angle of the $RZ(\pi/4)$ gate in Eq. (16). The tests were performed on the qubit 0, and we anticipated that the results would vary depending on the selected qubit, as shown in the previous report[3].

The first set of tests of the $RX(\pi/2)$ gate demonstrated decent fidelity and coherence of the gate. The measured

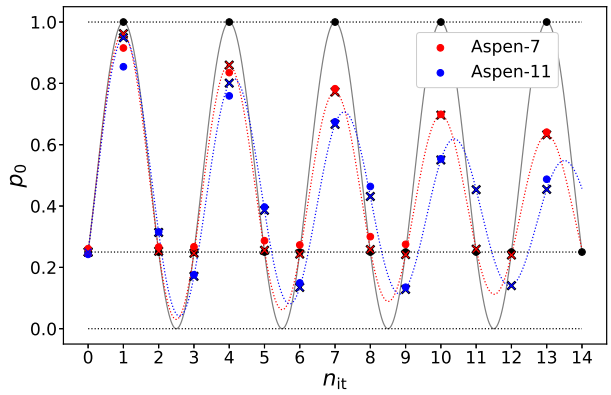

FIG. 6. 4-level GS on Aspen-7 (red) vs Aspen-11 (blue) for $\omega = 0$. Solid dots: experimental data; crosses: analytical fit using Eq. (15); dotted line: analytical function from Eq. (15) with fitted parameters.

over-rotation angle was about $\pi/10$ after 70 gate repetitions, resulting to the error/gate value $\Delta\phi_2 = -0.0034$ computed by fitting the data with the expression from Eq. (17). The errors of the $RX(\pi/4)$ gate were more pronounced with the error/gate value $\Delta\phi_4 = -0.0083$, which is slightly greater than expected $2\Delta\phi_2$, since two $RX(\pi/2)$ gates are involved in one $RX(\pi/4)$ gate. The data points for the $RX(\pi/4)$ gate test were collected at certain values of $n_{\mathtt{it}}$ that are spread apart in order to improve the accuracy of the fitting.

It is worth exploring possible strategies to mitigate over-rotation. In circuits consisting of hundreds of native gates, a relatively small portion typically involves 2-qubit entanglement gates (e.g., CZ or $CNOT$), while the majority are single-qubit RX/RZ gates. Based on our experiments, we observe a phase shift of a fraction of π due to the single-qubit gates alone. If there is a straightforward method to correct for this phase shift, it has the potential to significantly enhance the fidelity of quantum algorithms.

The strategy implemented here is centered on the notion that the $\pi/4$ rotation is achieved by the $RZ(\pi/4)$ gate in Eq. (16). Despite the fact that the RZ gate is virtual or software-based (not physically executed by the hardware), adjusting its angle can influence the angle of the resulting $RX(\pi/4)$ gate. The outcomes of these corrections are illustrated in Fig. (7). The uncorrected raw data of the repeating $RX(\pi/4)$ gate is shown with solid red dots. An oscillating red curve corresponds to the fitting function from Eq. (17). Disregarding the slow decay of the amplitude of oscillations, it is clear that the data is out of phase from the analytical prediction shown with the blue curve. The set of green points is the experimental data collected after the phase correction of the RZ gate, namely, changing the angle $\pi/4 \to \pi/4 + 0.0083$. While the decay rate (is similar to that of the shrinking red curve) is the same as before since it is determined by decoherent processes, the phase of the data points is in close agreement with the analytical curve. In fact, the angle of over-rotation after the correction is only $\Delta\phi_{cor} = 3.758 \cdot 10^{-5}$, which is 220 times smaller than that for the uncorrected data. In principle, this approach can be generalized to the case of arbitrary circuits, once the coherent errors of 2-qubit gates are thoroughly investigated.

While we do not provide values of the over-rotation angles of other qubits on the chip, it is worth mentioning that these values vary noticeably from qubit to qubit. This fact suggests that a possible cause of this error might be related to crosstalk, as the qubit frequency has a slight dependency on the state of its neighbors. As a result, a miscalibrated frequency manifests as the RZ gate error. The magnitude of this error is small, yet potentially is in line with the observed values.

VI. CZ PROTOCOL

Generally, the fidelity of 2-qubit entanglement gates is considerably lower than that of single-qubit gates. This trend is evident in the Rigetti platform, as it can be verified from the device calibration page[11]. Although the total number of entanglement gates is considerably smaller than the number of single qubit gates, the reduced fidelity of the former can have a substantial impact on overall hardware performance. In this analysis, we specifically concentrate on the CZ gate, a widely used native gate in Rigetti's quantum computing systems.

Distinguishing coherent and decoherent errors of a 2-qubit gate is a more intricate task than for a single qubit gate. Ideally, a comprehensive process tomography or even gate set tomography is required to obtain a detailed understanding of the CZ gate. However, due to limitations in QPU allocation, we had to devise simpler yet informative tests and procedures to assess qubit pair capabilities and gate performance. One intuitive approach involves measuring the decay rate of an excited state as a function of the CZ gate repetitions $n_{\mathtt{it}}$, similarly as it was described in Sec. (V). While such an approach might not be a good tool to detect phase errors, it targets the coherence time of the qubits, which is typically the limiting factor on the CZ gate fidelity. The setup of this benchmark is the following: (i) the system is prepared in the excited state $|01\rangle$, $|10\rangle$, or $|11\rangle$; (ii) the CZ gate is applied $n_{\mathtt{it}}$ times; (iii) the measurements are performed. Since the CZ gate changes phases only, the population density should not change as the gate applied. Due to decoherence, the population of the excited state decays and one can assume a simple exponential dependence $p_\omega \propto \exp(-n_{\mathtt{it}} T_{CZ}/T_1)$, where T_{CZ} is the length of the CZ control pulse that depends on the qubit but usually of the order $T_{CZ} = 200\ ns$ according to Rigetti data, and T_1 is the decoherence time of the excited qubit. Fig. (8) demonstrates the results of the CZ gate benchmarking performed on the qubit pair (0, 1). The x-axis is plotted using a logarithmic scale for better visibility of the tran-

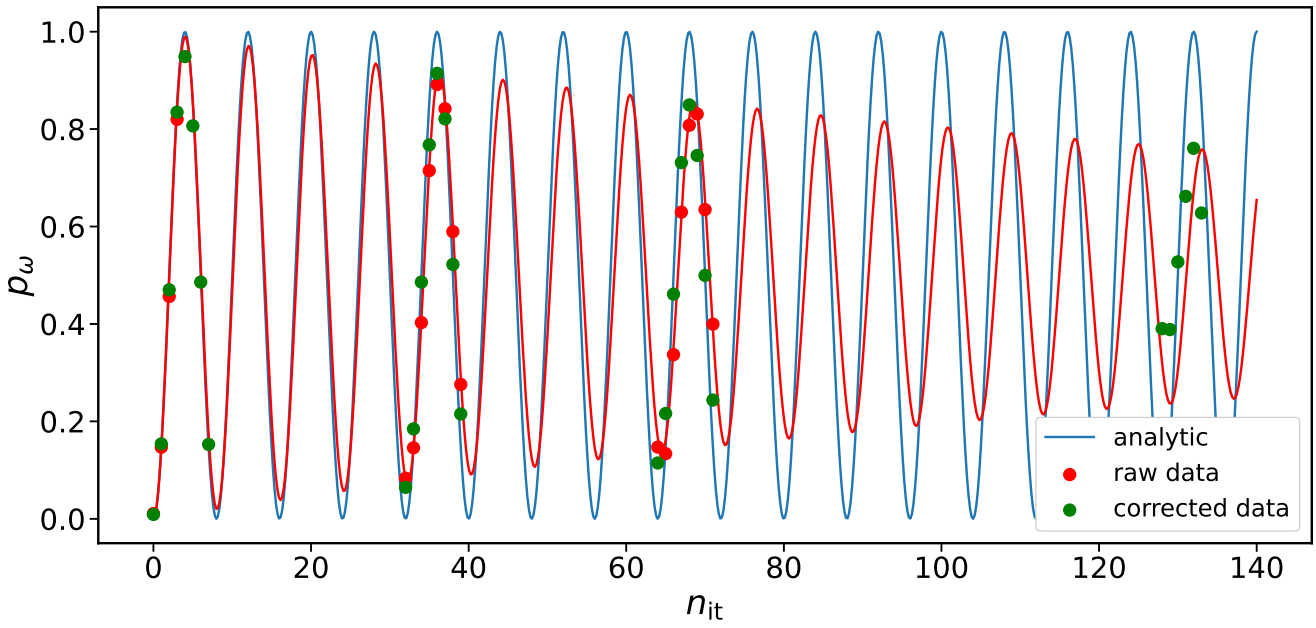

FIG. 7. Results of the $RX(\pi/4)$ gate tests conducted on the Aspen-11 QPU by repeating the gate multiple times. Red dots: measured experimental data points; red curve: fitting function from Eq. (17); green dots: experimental data points with corrected $RZ(\pi/4)$ gate in Eq. (16); blue solid curve: the analytic solution for the rotation of the qubit.

sient region. The solid points in the figure correspond to the measured data and the vertical lines represent measurement errors obtained according to the methodology described in Sec. (III). The bottom plot (figure (b)) shows an abnormally large standard deviation of the data, thus, it signals of the presence of hardware issues. Even disregarding this fact, the data imply that each of the two qubits within a selected pair can behave completely differently. Indeed, when approximated with exponential law, the decay rate of the qubit 0 is 0.125/0.00446=28 times greater than of the qubit 1. However, the ratio of reported T_1 times of the qubits in the device calibration page was only 3.5. Even more interesting behavior can be observed when both qubits are initially excited $|11\rangle$. While one of the qubits decays rapidly, the other qubit is still excited, which leads to the growth of the $|10\rangle$ state and quick collapse of the $|11\rangle$ state.

The unexpected behavior of the CZ gate and rapid excited state decay suggest that imperfections of the gate could be an important factor contributing to the poor overall performance of the Aspen-11 platform, particularly when the qubit pair (0, 1) is selected. Even without a detailed study of coherent errors, filtering out poorly performing qubit pairs can be achieved by conducting similar CZ gate tests on all qubit pairs on the QPU. Due to limited availability of the QPU allocation, conducting a comprehensive scan of fidelity decay for the CZ gate over a wide range of repetition numbers, as shown in Fig. (8), was unfeasible. Instead, we tested short and long circuits only, which comprise 10 and 100 repetitions of the CZ gate respectively. We define this procedure as

$Q_{\mathrm{CZ\_S}}$	$Q_{\mathrm{CZ\_L}}$	Q_{CZ}
(11, 26) 0.98890	(11, 26) 0.99543	(11, 26) 0.99216
(25, 26) 0.98528	(26, 27) 0.99499	(26, 27) 0.98998
(26, 27) 0.98499	(25, 26) 0.99411	(25, 26) 0.98969
(21, 36) 0.98130	(36, 37) 0.99299	(21, 36) 0.98687
(10, 17) 0.98074	(21, 36) 0.99246	(31, 46) 0.98625

TABLE III. Best qubit pairs of Aspen-11 based on three CZ protocols. Colors are used to highlight the same qubit pairs in different versions of the protocols.

the CZ *protocol*, where every pair gets 4 scores (denoted as Q): for two initially excited states and for two circuit lengths. The score is calculated as follows. For a given qubit pair (k, m) the system is initially prepared in either $|01\rangle$ or $|10\rangle$ state. Then, the CZ gate is repeated either $n_{\mathrm{it}} = 10$ or $n_{\mathrm{it}} = 100$ times and the population of the excited state is measured. The obtained numbers are multiplied for a given value of n_{it} and the result is raised to the power $1/(2n_{\mathrm{it}})$ in order to normalize the value to the number of CZ gates applied. We use labels "S" and "L" for the short and long versions of the protocol ($n_{\mathrm{it}} = 10$ and $n_{\mathrm{it}} = 100$), thus, for example, the score denoted as $Q_{\mathrm{CZ\_S}}(k, m)$ roughly represents the following statement: "The average over the initial state and the number of repetitions $n_{\mathrm{it}} = 10$ fidelity drop per CZ gate, obtained from the short version of the CZ protocol applied to the qubit pair (k, m)".

Examples of the best 5 qubit pairs selected by the CZ protocols are shown in Table (III). We also introduced a

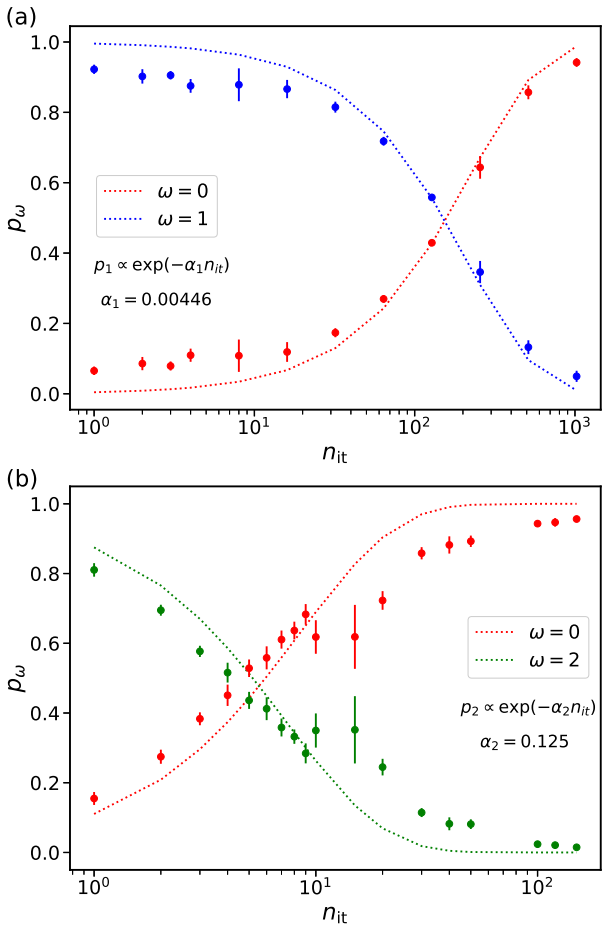

FIG. 8. Decay of the excited state population as a function of the number of the CZ gate repetition n_{it}: (a) for the excited state $\omega = 1$; (b) for $\omega = 2$. The decay rate α is calculated using the exponential formula $p_\omega \propto \exp(-\alpha_\omega n_{\text{it}})$.

global CZ protocol with scores labeled as $Q_{\text{CZ}}(k, m)$ and calculated as

$$Q_{\text{CZ}}(k,m) = \left[Q_{\text{CZ\_S}}(k,m) \cdot Q_{\text{CZ\_L}}(k,m)\right]^{1/2}. \qquad (18)$$

As it follows from Table (III), the best pairs according to one protocol might be not exactly the best according to another, for instance, pairs (25, 26) and (26, 27). This behavior is anticipated, as the decay rate could vary strongly depending on n_{it} in a way demonstrated in Fig. (8). Nevertheless, it is almost never the case when a good qubit pair according to one version of the CZ protocol performs poorly according to another version, which proves the relative robustness of the implemented procedure. The efficiency of the CZ protocol was shown in Sec. (IV) where the results from qubit pairs (0, 1) and (11, 26) were discussed.

It is also worth mentioning that the long version of the protocol maybe less informative, since 100 CZ gates might be over the limit of decoherence time (or gate depth) for some of the qubits on the chip. Clearly, when the system becomes nearly decoherent, any systematic measurement error, or noise, or accidental spike of the measured population density becomes dominant and eventually amplified by this type of a protocol that can potentially spoil the results. Unfortunately, we recognized this issue after the Aspen-11 lattice had already been retired, so that we could not recapture new data with fewer repetitions of the CZ gate.

VII. GROVER'S SEARCH PROTOCOLS

The CZ protocol was shown to be a reasonable tool for qubit pair selection. However, as it was pointed out before, it does not necessarily detect coherent errors (or provide quantitative measure of them). Indeed, assuming that the CZ gate is not a diagonal matrix with elements $(1, 1, 1, -1)$ but with $(1, 1, 1, i)$ instead, there is no way to determine the change in phase by applying the gate multiple times. At the same time, such a large phase change will ruin almost any algorithm by destroying all of the expected phases completely. In order to address these issues, we also developed a set of Grover's search protocols (or *GS protocols*). These protocols loop over all the qubit pairs in the same way that the CZ protocol does, except that they run either the 4-level or 3-level GS algorithm and assign scores based on the measured state population. When designing the GS protocols, it was important to choose the number of Grover's iterations reasonably. Having too few iterations yields better fidelity, however the results are strongly affected by SPAM errors, and this complicates analysis of gate errors. On the other hand, having too many iterations drives the system to almost decoherent state virtually eliminating all the purpose of using the protocol. Furthermore, normalization of the protocol scores in the way it was done for the CZ protocol reduces the sensitivity of the protocols.

Having these constraints considered, we designed 3-level and 4-level versions of the GS protocol with short "S" and long "L" subversions, denoted as GS3L_S, GS3L_L, GS4L_S and GS4L_L. The 3-level GS3L_S includes 2 Grover's iterations, so that the desired population of the Oracle marked state equals to $p_\omega(n_{\text{it}} = 2) = 0.0041$, while GS3L_L comprises 6 iterations with the analytical expected value $p_\omega(n_{\text{it}} = 6) = 0.9784$. GS4L_S and GS4L_L include 4 and 7 iterations respectively with $p_\omega = 1.0$ in both cases. Using the expression for the CZ gate count in both cases given in Eq. (13), the normalizations for the GS protocol scores are

$$\begin{cases} Q_{\text{GS3L\_S}} = (q_0 q_1 q_2)^{1/27}\big|_{n_{\text{it}}=2} \\ Q_{\text{GS3L\_L}} = (q_0 q_1 q_2)^{1/63}\big|_{n_{\text{it}}=6} \\ Q_{\text{GS4L\_S}} = (q_0 q_1 q_2 q_3)^{1/32}\big|_{n_{\text{it}}=4} \\ Q_{\text{GS4L\_L}} = (q_0 q_1 q_2 q_3)^{1/56}\big|_{n_{\text{it}}=7} \end{cases} \qquad (19)$$

Here, q_k is the measure of fidelity decay for k-th Marked Oracle state, namely, $q_k = 1 - |p_k - p_{k,a}|$, where p_k is

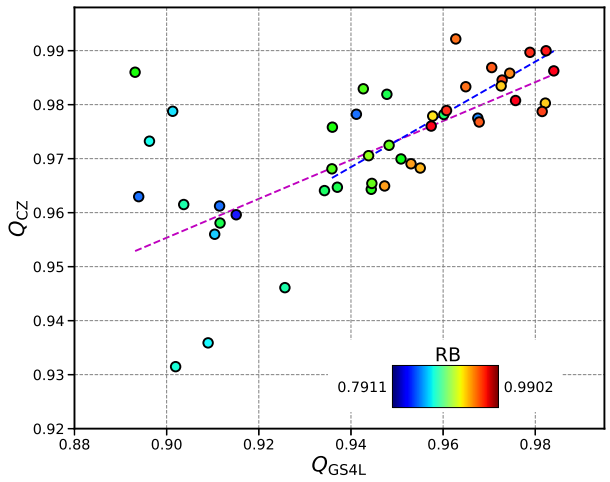

FIG. 9. Correlation between the GS4L and CZ protocols with the RB score depicted with a color map. The results show scores for all possible qubit pairs on the Aspen-11 QPU. The magenta and blue dashed lines shows the Theil–Sen estimator for all qubit pairs and for the best 23 pairs according to RB scores respectively. The slopes are given by $Q_{\texttt{CZ}} = 0.36 \cdot Q_{\texttt{GS4L}} + 0.63$ for magenta, and $Q_{\texttt{CZ}} = 0.49 \cdot Q_{\texttt{GS4L}} + 0.51$ for blue lines.

the measured population of the state and $p_{k,a}$ is the analytical value. For example, in the case the 4-level GS protocol, all $p_{k,a} = 1$, thus, $q_k = p_k$. Using a definition that is similar to the global CZ protocol, we now introduce global GS protocols (GS3L and GS4L) via

$$\begin{cases} Q_{\texttt{GS3L}}(k,m) = [Q_{\texttt{GS3L\_S}}(k,m) \cdot Q_{\texttt{GS3L\_L}}(k,m)]^{1/2}, \\ Q_{\texttt{GS4L}}(k,m) = [Q_{\texttt{GS4L\_S}}(k,m) \cdot Q_{\texttt{GS4L\_L}}(k,m)]^{1/2}. \end{cases} \tag{20}$$

Finally, scores from all the hardware protocols are combined in one "Best Hardware" protocol (*BH protocol*), where scores are calculated as a geometric mean of all the protocols

$$Q_{\texttt{BH}}(k,m) = [Q_{\texttt{CZ}}(k,m) \cdot Q_{\texttt{GS3L}}(k,m) \cdot Q_{\texttt{GS4L}}(k,m)]^{1/3}. \tag{21}$$

When some of the data from some of the protocols were missing (see Sec. (VII A)), Eq. (21) was adjusted in a way that the scores from available hardware protocols were weighted equally.

A. Tests of Aspen-11

The GS protocols were first applied to the 38-qubit Aspen-11 QPU in 2022. In the early stages of our study, GS4L_L was the only hardware protocol that was cross-verified against the CZ protocol. The protocols were designed not only for benchmarking Rigetti hardware and selecting the best qubit pair but also for comparison to Rigetti's calibration data, including qubit T_1 and T_2 times and the results of randomized benchmarking (RB) for the CZ gate on all available qubit pairs. The goal was to identify correlations between the hardware protocols and provided calibration data. Such correlations could greatly improve the efficiency of QPU allocation usage for the best qubit pair selection. Specifically, running the CZ protocol for all pairs on the Aspen-11 chip with $n_b = 5$ and $N_s = 10^3$ takes about 22 minutes, which is the minimum for a production run. The GS4L_L protocol takes less time, but for larger chips like 80-qubit Aspen-M-2 or Aspen-M-3, completing the protocols even in one hour becomes impossible. If calibration data helps preselect the best candidates, it could significantly enhance the selection process.

Both CZ and GS protocols underwent cross-verification against $T_{1,2}$ and RB data. To determine the correlation with the T_1 decoherence time, we compared the measured decay rate scores of the CZ/GS protocols with the estimated decay rate based on the T_1 time of the qubits. Unfortunately, no correlation between CZ/GS scores and the $T_{1,2}$ decay rates was observed, leading to the exclusion of $T_{1,2}$ data from further consideration. The comparison of the GS4L_L, CZ and RB scores is shown in Fig. (9). A "jet" color map (blue to red transition) is used to represent the third dimension on the plot. With some exceptions, all three metrics are more or less consistent with each other, which is especially important for the data points of highers scores (upper-left corner of Fig. (9)). Poor correlation between the GS and CZ protocols for low score pairs can be explained by strong coherent and decoherent errors that are treated differently by GS vs. CZ circuits. Nevertheless, solely relying on the RB data is insufficient for the best qubit pair selection. To be more specific, consider the qubit pair (14, 15), which has a better RB score than the pair (26, 27). According to the BH protocol (which is the geometric mean of the CZ and GS4L in this particular realization) however, the scores of the pairs are 0.9782 and 0.9861 respectively. For a small circuit consisting of 20 CZ gates, the fidelity of the pair (26, 27) is estimated to be 0.75, while for the pair (14, 15) it is 0.64, which leads to a significantly different conclusion.

B. Tests of Aspen-M-2

With the introduction of the new Rigetti QPUs, Aspen-M-2 (already retired at the moment of writing of this manuscript) and Aspen-M-3, we employed all aforementioned hardware protocols. As a result, the run time of the protocols significantly increased, especially considering the expansion of the number of qubits and qubit pairs on a chip from 38 to 80. With typical circuit parameters, such as $N_s = 1000$ number of shots and $n_b = 5$ number of statistical batches, it was possible to scan only 20-25 qubit pairs in 40 minutes. Despite these challenges, a substantial correlation between the hardware protocols

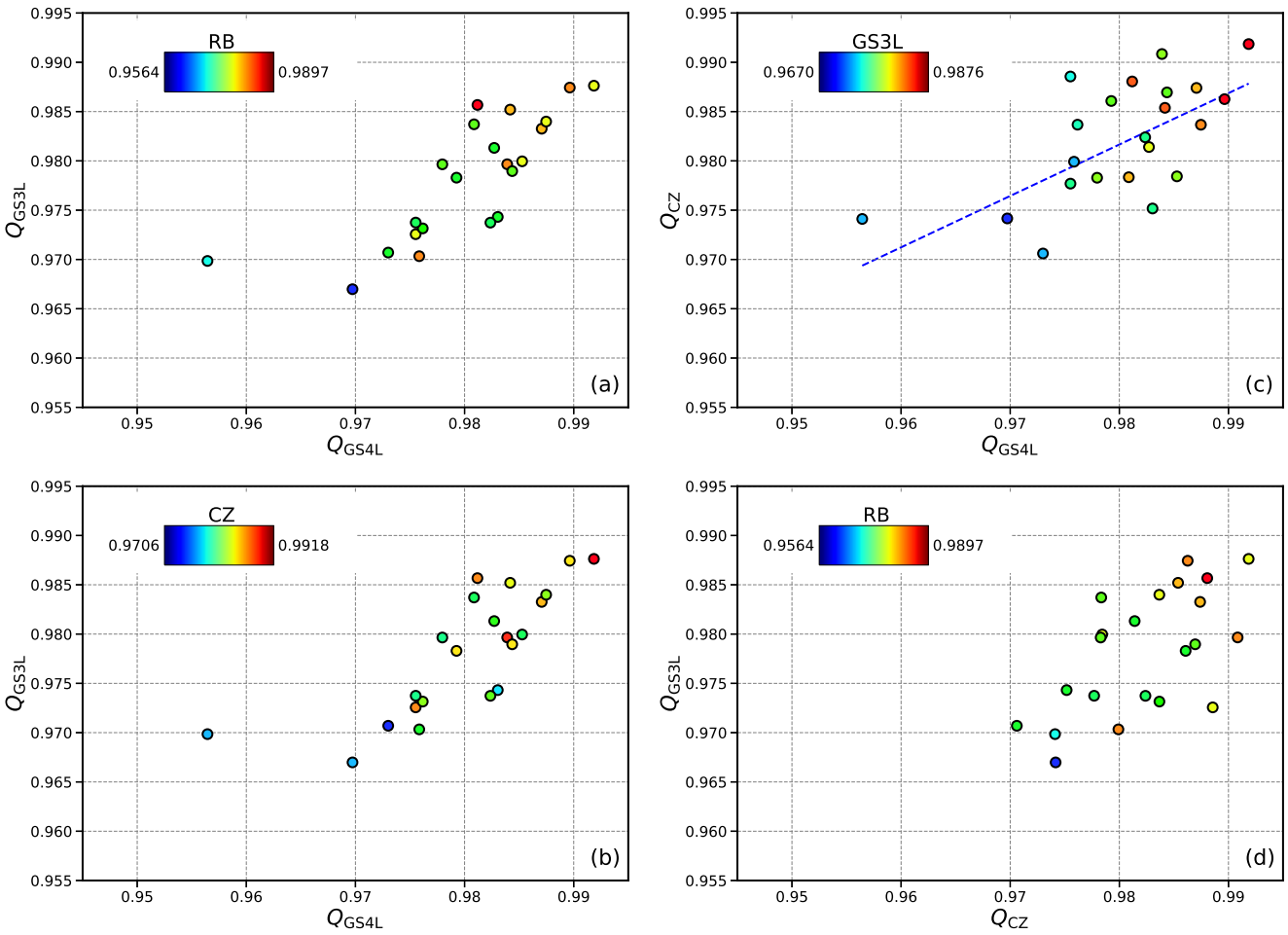

FIG. 10. Correlation between different hardware protocols and RB scores performed on Aspen-M-2. Figure (a): CG4L - x-axis, GS3L - y-axis, RB - color map. Figure (b): CG4L - x-axis, GS3L - y-axis, CZ - color map. Figure (c): CG4L - x-axis, CZ - y-axis, GS3L - color map. Figure (d): CZ - x-axis, GS3L - y-axis, RB - color map. Color map: red - high score, blue - low score. The blue line in figure (c) shows the Theil–Sen estimator with $Q_{\text{CZ}} = 0.52 \cdot Q_{\text{GS4L}} + 0.47$. The deliberate inclusion of additional spacing in the figures (a-d) is intended to accurately align with the axes of Aspen-M-2 and Aspen-M-3 data plots (see Fig. (11)), facilitating enhanced comparability.

and RB data, as discussed in Sec. (VII A), validated the use of RB data as a pre-selection tool for Aspen-M-2 and Aspen-M-3. Instead of looping over all pairs on the chip, the combined hardware protocol tested those qubit pairs with the best RB score first. A possible speed-up of the protocol could be realized by playing gates in parallel using a next-nearest-neighbor as an insulator. However, to ensure the complete elimination of any potential crosstalk between the tested pairs, we performed our tests in serial. Nevertheless, parallel protocol execution persists as a major improvement and it is planned to be tested in future works.

Fig. (10) demonstrates results of all protocols and the RB scores shown using a color scheme, similarly how it was done in Fig. (9). Figures (a) and (b) show the correlation between the two different Grover's search protocols. While the correlation between GS4L and GS3L data is decent, both RB and CZ metrics (depicted with color) are not in good agreement with the results of the GS protocols. This observation might signal of the presence of coherent errors that are better captured by the GS protocols rather than repetition of the CZ gate. Similar conclusions could be drawn from the comparison of the CZ protocol data with GS4L (figure (c)) and GS3L (figure (d)). This result can be verified in Table (IV) where the scores of top 8 pairs of all protocols are collected. The only pair that is the best according to all metrics is (16, 17).

Another aspect we investigated was the temporal stability of qubit pairs. Given that quantum hardware undergoes regular retuning, understanding the extent of variations in the properties of qubits and qubit pairs over time is crucial. We categorized time stability into short-term and long-term. Short-term stability is particularly

Protocol	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Rank 6	Rank 7	Rank 8
Data collected: 23 March 2023								
RB	(122, 135) 0.989658	(10, 17) 0.985085	(10, 113) 0.984489	(102, 103) 0.983706	(1, 16) 0.983571	(110, 117) 0.983052	(31, 46) 0.979880	(16, 17) 0.979614
CZ_S	(16, 17) 0.990361	(10, 113) 0.988491	(110, 117) 0.984307	(122, 135) 0.983654	(101, 102) 0.982848	(31, 46) 0.982386	(126, 127) 0.980672	(10, 17) 0.979076
CZ_L	(31, 46) 0.994768	(1, 16) 0.993870	(10, 17) 0.993504	(16, 17) 0.993339	(10, 113) 0.993195	(122, 135) 0.992478	(126, 127) 0.991508	(101, 102) 0.991066
CZ	(16, 17) 0.991849	(10, 113) 0.990840	(31, 46) 0.988558	(122, 135) 0.988056	(110, 117) 0.987402	(101, 102) 0.986949	(10, 17) 0.986264	(126, 127) 0.986075
GS4L_S	(16, 17) 0.992429	(10, 17) 0.992009	(141, 142) 0.990533	(110, 117) 0.988827	(130, 131) 0.986827	(10, 11) 0.986542	(10, 113) 0.986263	(110, 111) 0.986082
GS4L_L	(16, 17) 0.991274	(10, 17) 0.987278	(110, 117) 0.985313	(1, 16) 0.985097	(110, 111) 0.984471	(141, 142) 0.984424	(10, 11) 0.984155	(101, 102) 0.982844
GS4L	(16, 17) 0.991851	(10, 17) 0.989641	(141, 142) 0.987474	(110, 117) 0.987068	(10, 11) 0.985348	(110, 111) 0.985276	(101, 102) 0.984365	(1, 16) 0.984179
GS3L_S	(16, 17) 0.993950	(10, 17) 0.993085	(122, 135) 0.992769	(1, 16) 0.991349	(141, 142) 0.991216	(23, 24) 0.990834	(110, 117) 0.990381	(101, 102) 0.990076
GS3L_L	(10, 17) 0.981819	(16, 17) 0.981343	(1, 16) 0.979090	(45, 46) 0.978736	(122, 135) 0.978630	(141, 142) 0.976816	(110, 117) 0.976195	(10, 113) 0.974071
GS3L	(16, 17) 0.987626	(10, 17) 0.987436	(122, 135) 0.985674	(1, 16) 0.985200	(141, 142) 0.983990	(45, 46) 0.983710	(110, 117) 0.983262	(23, 24) 0.981319
BH	(16, 17) 0.990440	(10, 17) 0.987779	(110, 117) 0.985909	(141, 142) 0.985042	(122, 135) 0.984966	(1, 16) 0.984920	(10, 113) 0.984791	(101, 102) 0.983420
Data collected: 28 March 2023								
RB	(10, 113) 0.988811	(101, 102) 0.987497	(1, 16) 0.986307	(122, 135) 0.982492	(112, 113) 0.981391	(10, 17) 0.980089	(122, 123) 0.979762	(23, 24) 0.978835
CZ_S	(122, 135) 0.985337	(10, 113) 0.985076	(132, 145) 0.982705	(130, 131) 0.981310	(1, 16) 0.981306	(16, 17) 0.979784	(10, 17) 0.979529	(110, 117) 0.978814
CZ_L	(1, 16) 0.994403	(31, 46) 0.993973	(10, 17) 0.993334	(122, 135) 0.993104	(10, 113) 0.992995	(16, 17) 0.992834	(23, 24) 0.991670	(10, 11) 0.991512
CZ	(122, 135) 0.989213	(10, 113) 0.989027	(1, 16) 0.987833	(132, 145) 0.986464	(10, 17) 0.986407	(31, 46) 0.986333	(130, 131) 0.986326	(16, 17) 0.986287
GS4L_S	(141, 142) 0.988006	(101, 102) 0.986931	(1, 16) 0.986209	(122, 135) 0.985623	(16, 17) 0.984698	(130, 131) 0.984669	(10, 17) 0.983158	(110, 117) 0.982077
GS4L_L	(141, 142) 0.986146	(1, 16) 0.985680	(16, 17) 0.984009	(110, 117) 0.982717	(10, 17) 0.982271	(122, 135) 0.981938	(122, 123) 0.981359	(101, 102) 0.979774
GS4L	(141, 142) 0.987075	(1, 16) 0.985944	(16, 17) 0.984354	(122, 135) 0.983779	(101, 102) 0.983346	(10, 17) 0.982714	(110, 117) 0.982397	(122, 123) 0.981694
GS3L_S	(1, 16) 0.992735	(16, 17) 0.991962	(122, 135) 0.991782	(141, 142) 0.991348	(122, 123) 0.989247	(10, 17) 0.989035	(23, 24) 0.988929	(110, 117) 0.988499
GS3L_L	(1, 16) 0.981567	(122, 135) 0.979367	(141, 142) 0.977064	(16, 17) 0.976629	(10, 113) 0.975083	(110, 117) 0.974797	(45, 46) 0.974604	(10, 17) 0.972823
GS3L	(1, 16) 0.987135	(122, 135) 0.985555	(16, 17) 0.984266	(141, 142) 0.984180	(110, 117) 0.981624	(10, 17) 0.980895	(10, 113) 0.980740	(122, 123) 0.980673
BH	(1, 16) 0.986970	(122, 135) 0.986179	(16, 17) 0.984968	(141, 142) 0.983927	(10, 17) 0.983336	(10, 113) 0.982978	(110, 117) 0.982724	(130, 131) 0.982643

TABLE IV. Collection of RB data and all hardware protocol scores of the Aspen-M-2 QPU obtained in March 2022. Red and green colors highlight scores of the two qubit pairs with the best BH score from 03/23/2023. Blue and magenta colors highlight scores of the two qubit pairs with the best BH score from 03/28/2023.

significant, as any production runs become futile if qubit fidelity experiences significant drifts within a time scale of an hour. However, in this case, we conducted experiments to examine long-term stability, and the results are presented in Table (IV). We ran all the protocols and recorded data for the top 8 qubit pairs identified by each protocol. These experiments were conducted with a 5-day interval, on 03/23/2023 and 03/28/2023. Qubit pairs were marked with four different colors (red, blue, green, magenta) based on their performance, with two pairs from each day of the experiment. Although some qubit pairs exhibited significant movement or were excluded from the top-performing list, the majority of the best pairs still maintained their positions in the top 5 according to each protocol. Long-term QPU stability could be of possible interest for optimizing qubit pair selection.

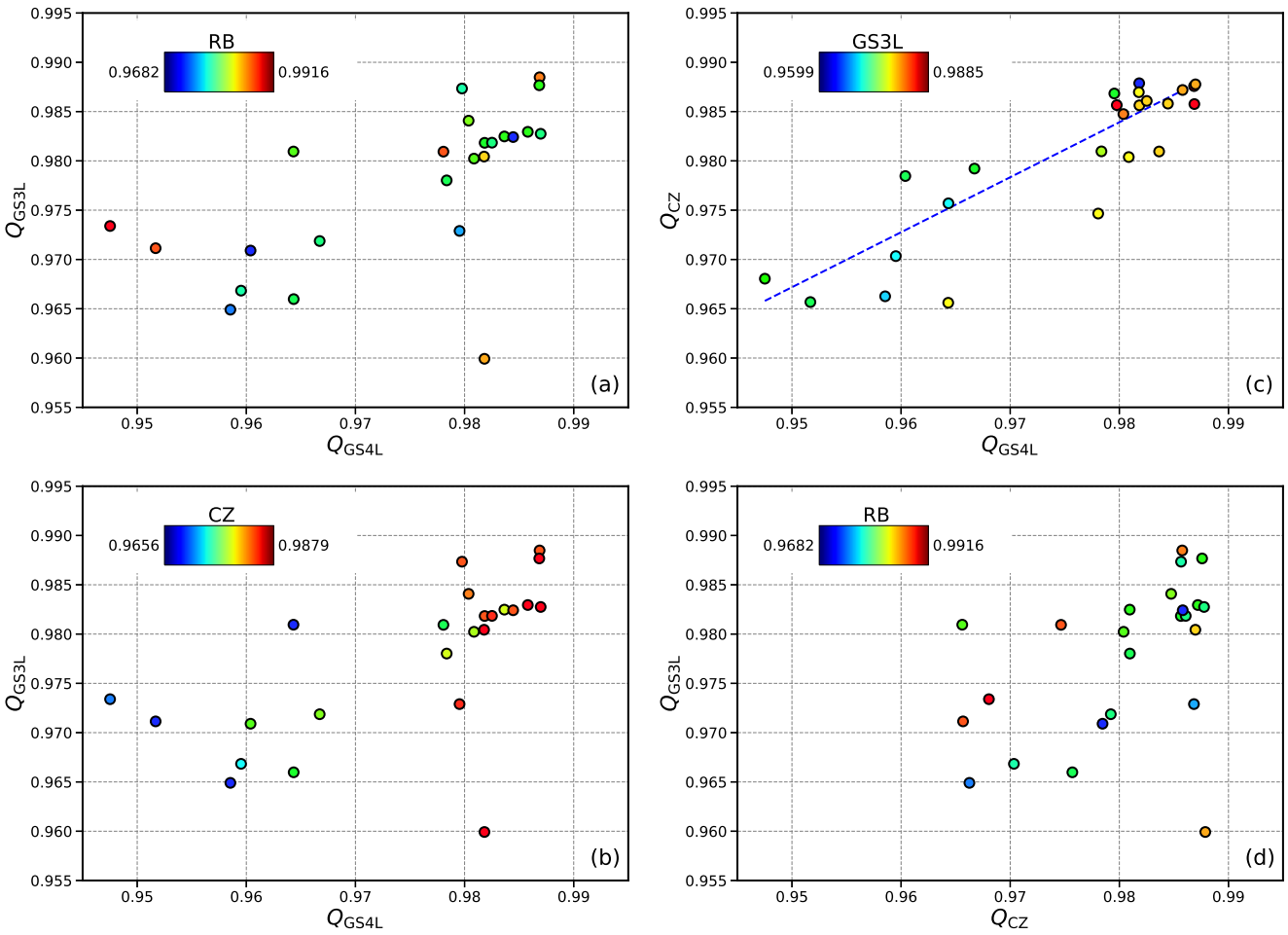

FIG. 11. Correlation between different hardware protocols and RB scores performed on Aspen-M-3. Figure (a): CG4L - x-axis, GS3L - y-axis, RB - color map. Figure (b): CG4L - x-axis, GS3L - y-axis, CZ - color map. Figure (c): CG4L - x-axis, CZ - y-axis, GS3L - color map. Figure (d): CZ - x-axis, GS3L - y-axis, RB - color map. Color map: red - high score, blue - low score. The blue line in figure (c) shows the Theil–Sen estimator with $Q_{\texttt{CZ}} = 0.56 \cdot Q_{\texttt{GS4L}} + 0.44$.

Instead of relying solely on RB data for pre-selecting the best qubit pairs, we could perform one scan of all available pairs once. Assuming qubits change their properties slowly, we could then run the protocols using the best pairs identified in the previous hardware protocol run.

C. Tests of Aspen-M-3

The set of tests conducted on the Aspen-M-3 chip was similar to those performed on Aspen-M-2. The results are shown in Fig. (11), where the subplots, axes (including ranges of values in both x- and y- directions), and color maps are identical as in Fig. (10). In general, the behavior of the Aspen-M-3 QPU is slightly worse than Aspen-M-2, as the characteristic spread of the data is greater, which implies worse correlation between different metrics. The only noticeable exception is a decent correlation between all three hardware protocols for 10-15 best pairs that is shown as a cluster of points in the upper right corner of figure (b) where almost all the data points have red color. The correlation between the RB data and hardware protocols is still below expected and this raises concerns about the pre-selection method used.

Short-term temporal qubit stability, introduced and briefly discussed in Sec. (VII B) was tested on the Aspen-M-3 chip. After all pre-selected qubit pairs were processed by the protocols, we could immediately run required post-processing tools to find the best and the worst pairs. We then ran the 4-level Grover's search algorithm on the best and the worst pairs trying to reveal the nature of their performance using Eq. (15). While overall the pairs behaved accordingly with the assigned by the protocols scores, there was one exception, namely the qubit pair (120, 121). It received a mediocre score in all hardware protocols, yet several minutes after it performed exceptionally well in the Grover's search test with $p_\omega = \{0.923, 0.939, 0.895, 0.956\}$ for GS4L_S

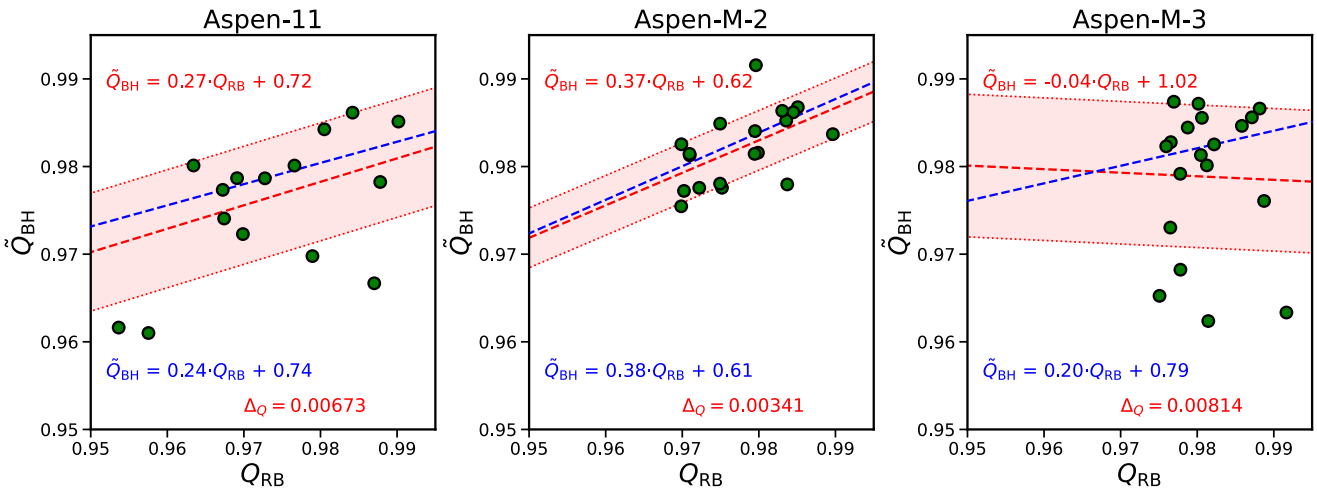

FIG. 12. Comparison of the RB and BH data with a linear regression model (red dashed) and the robust Theil–Sen estimator (blue dashed). The pink domain shows the bounds of the data within standard deviation $\pm\Delta_Q$ of linear regression. Here, $\tilde{Q}_{\rm BH}$ denotes scores of the best hardware protocol based on the GS4L and CZ protocols only.

and $p_\omega = \{0.900, 0.818, 0.812, 0.876\}$ for GS4L_L circuits, which is probably the best performance of the Rigetti hardware ever registered throughout our three year experience with this platform. This appears to indicate that issues with hardware stability still remain for Aspen-M-3.

VIII. CONCLUSION

With the evolution of quantum hardware and the increase in the total number of qubits, managing qubit controls, addressing crosstalk issues, and optimizing circuits to meet emerging constraints have become increasingly challenging. One of these challenges is the freedom of selecting qubit pairs. This report thoroughly addressed this issue by implementing various hardware protocols on the Rigetti Aspen-11, Aspen-M-2, and Aspen-M-3 QPUs. The three protocols consisted of 3- and 4-level Grover's search protocols, along with the CZ protocol. Given that each protocol had both short and long versions, they proved to be effective tools for identifying various hardware errors and selecting the best qubit pairs with minimal susceptibility to hardware issues.

The extensive volume of generated data, obtained through the exploration of available qubit pairs using the devised protocols, can be condensed and summarized effectively through the best hardware (BH) scores. These scores represent the geometric mean of all scores from various protocols, assuming equal weights for each. In essence, these scores provide a rough indication of the fidelity drop resulting from a single application of the CZ gate. One of the main goals of the present research was to verify whether the RB data can be used instead of the hardware protocols as a qubit pair selection tool. Fig. (12) summarizes our findings by showing comparison of the BH scores with the RB data across all Rigetti lattices we tested in 2022-2023. Surprisingly, the newest Aspen-M-3 chip shows the worst performance in terms of the correlation between the RB and BH data as well as the standard deviation spread. Unfortunately, statistics of approximately 20 data points is quite scarce, therefore linear regression does not provide reasonable estimates for the slope and intercept values. This is evident in the Aspen-M-3 plot, where the observed negative slope is deemed entirely unphysical. To address this issue, we employed the robust Theil–Sen estimator, which not only aligns well with linear regression for Aspen-M-2 but also resolves the problem encountered with Aspen-M-3. Such an approach provides the median value of the slopes κ over all pairs of data points, thus, the variability in the median slope value across different QPUs can serve as a metric to assess the robustness of correlation between various protocols. For example, Fig. (12) shows that for the BH-RB plots, the value κ varies from 0.2 to 0.38 among all three Aspen chips we tested. For comparison, the same test done for the CZ-GS4L set of plots (see Figures (9-11)) yields much smaller variation of κ, namely from 0.49 to 0.56, revealing better correlation between hardware protocols than from hardware protocols and RB scores. In any case, the key takeaway is that relying solely on RB data is not recommended for selecting the best qubit pair.

As part of our second focus, we conducted an analysis of hardware errors with the aim of decomposing them into coherent and decoherent components. Our findings revealed that coherent errors can be mitigated, as discussed in Sec. (V), by applying phase corrections to the RZ gate. A potential next step could involve employing process tomography to obtain results for the CZ gate, enabling the identification of methods to correct its co-

herent errors as well. Taking the analysis a step further, an even more challenging task is to improve the hardware protocols. The goal would be not only to distinguish between good and bad qubit pairs but also to provide insights into the level of coherent errors if they are present. This would contribute to a more detailed understanding of the quantum hardware's performance and pave the way for targeted improvements.

IX. ACKNOWLEDGMENTS

This work was performed under the auspices of U.S. Department of Energy (DOE) by Lawrence Livermore National Laboratory (LLNL) under Contract No. DE-AC52-07NA27344 and was supported by the DOE Office of Fusion Energy Sciences under the "Quantum Leap for Fusion Energy Sciences" project SCW1680 and the "Recreating Plasma Dynamics on an Intermediate-Scale Quantum Processor" project SCW1736. Quantum compute time was provided by Rigetti and by the Quantum Computing User Program at the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC05-00OR22725. The authors would like to express their gratitude to the Rigetti Computing team, acknowledging them separately for providing access to Aspen-11, Aspen-M-2, and Aspen-M-3 quantum processors, as well as the valuable discussions and guidance.

REFERENCES

[1] L. K. Grover, "A fast quantum mechanical algorithm for database search," (1996), arXiv:9605043 [quant-ph].
[2] Rigetti compiler can be found at https://pyquil-docs.rigetti.com/en/stable/compiler.html.
[3] V. Geyko, "Using grover's search algorithm to characterize the rigetti quantum computing platform," Tech. Rep. LLNL-TR-819320 1025951 (Lawrence Livermore National Laboratory, 2021).
[4] E. Nielsen, J. K. Gamble, K. Rudinger, T. Scholten, K. Young, and R. Blume-Kohout, Quantum **5**, 557 (2021).
[5] M. Paris and J. Řeháček, *Quantum State Estimation*, 1st ed., Lecture Notes in Physics (Springer Berlin, Heidelberg, 2004) pp. XIII, 520, published: 18 August 2004.
[6] H. O. Kunz, IEEE Trans. Comps. **C-28**, 267 (1979).
[7] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information* (Cambridge university press, 2010).
[8] PRAGMA PRESERVE_BLOCK https://pyquil-docs.rigetti.com/en/1.9/compiler.html.
[9] S. Kimmel, G. H. Low, and T. J. Yoder, Phys. Rev. A **92**, 062315 (2015).
[10] A. M. Meier, K. A. Burkhardt, B. J. McMahon, and C. D. Herold, Physical Review A **100** (2019), 10.1103/physreva.100.052106.
[11] Rigetti device calibration page https://qcs.rigetti.com/lattices.